

PATENT APPLICATION

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE  
BEFORE THE HONORABLE BOARD OF PATENT APPEALS AND INTERFERENCES

In re the Application of

Francine R. Chen et al.

Application No.: 10/626,875

Examiner: Tomasz Ponikiewski

Filed: July 25, 2003

Docket No.: A3053-US-NP  
XERZ 2 01564

Title: SYSTEMS AND METHODS FOR LINKED EVENT DETECTION

BRIEF ON APPEAL

Appeal from Group 2165

Mark S. Svat, Esq., Reg. No. 34,261  
FAY SHARPE LLP  
1100 Superior Avenue – Seventh Floor  
Cleveland, Ohio 44114-2579  
Telephone: (216) 861-5582  
Attorneys for Appellants

**CERTIFICATE OF TRANSMISSION**

I hereby certify that this correspondence (and any item referred to herein as being attached or enclosed) is (are) being transmitted to the USPTO by electronic transmission via the EFS Web on the date indicated below.

Date

*April 14, 2008*

Name:

*[Signature]*

TABLE OF CONTENTS

	<u>Page</u>
I. REAL PARTY IN INTEREST .....	1
II. RELATED APPEALS AND INTERFERENCES .....	2
III. STATUS OF CLAIMS .....	3
IV. STATUS OF AMENDMENTS .....	4
V. SUMMARY OF CLAIMED SUBJECT MATTER.....	5
VI. GROUNDS OF REJECTION TO BE REVIEWED ON APPEAL .....	12
VII. ARGUMENT .....	13
A. Claims 1, 20, 39, 58, and 81-84 Would Not Have Been Obvious Over Sundaresan in View of Pirolli and Further in View of Maybury ...	13
1. Claim 1 .....	13
a. Sundaresan Does Not Teach Use of Source Identification ..	13
b. Combination of Sundaresan and Pirolli Does Not Teach Determining Inter-Story Similarity Vectors for Source- Identified Training Stories.....	14
c. Sundaresan Does Not Teach Determining Link label Information for Source-Identified Training Stories .....	14
d. Maybury Does Not Teach Indicating Existence of Stories Linked to the Same Event .....	15
2. Claims 20, 39, 58 and 81-84 .....	16
3. Claim 2 .....	16
4. Claims 21, 40 and 59.....	18
B. B.Claim 77 Would Not Have Been Obvious Over Brown in View of Arend	19
VIII. CLAIMS APPENDIX .....	23
IX. EVIDENCE APPENDIX .....	39
X. RELATED PROCEEDINGS APPENDIX.....	40

I. REAL PARTY IN INTEREST

The real party in interest for this appeal and the present application is Palo Alto Research Center Inc. (3333 Coyote Hill Rd., Palo Alto, California 94304), by way of an Assignment recorded in the U.S. Patent and Trademark Office at Reel 014333, Frame 0512.

## II. RELATED APPEALS AND INTERFERENCES

There are no prior or pending appeals, interferences or judicial proceedings, known to Appellant, Appellant's representative, or the Assignee, that may be related to, or which will directly affect or be directly affected by or have a bearing upon the Board's decision in the pending appeal.

There is, however, a provisional rejection of claims 1, 20, 39, 58, and 81-84 on the ground of non-statutory obviousness-type double patenting over claims 1, 16, and 31-32 of co-pending Application No. 10/626,856.

III. STATUS OF CLAIMS

Claims 1-88 are on appeal.

Claims 1-88 are pending.

Claims 1-88 are rejected.

#### IV. STATUS OF AMENDMENTS

An Amendment After Final Rejection was filed on January 21, 2008. By an Advisory Action dated February 19, 2008, it was indicated that the requested amendments were not entered. Therefore, the claims presented for appeal are those as set forth prior to Applicants' Amendment After Final submitted January 21, 2008.

## V. SUMMARY OF CLAIMED SUBJECT MATTER

The invention of claim 1 is directed to a computer-implemented method of determining predictive models for a linked event detection system as shown in FIG. 2, including: determining source-identified training stories, determining inter-story similarity vectors in a memory **20** (FIG. 4) for at least one story-pair of the source-identified training stories, determining link label information for the at least one story-pair, and determining and storing at least one predictive model in the memory based on the inter-story similarity vectors and the link label information. The link label information indicates the existence of at least one link between a pair of stories in the source-identified training stories and that the linked source-identified stories are related to the same event.

A processor **15**, as shown in FIG. 4 of the specification, performs the step of determining source-identified training stories and the step of determining link label information as described on page 20, line 20 – page 21, line 3. The step for determining inter-story similarity vectors is performed by an inter-story similarity determining circuit **40** as described on page 21, line 28 – page 22, line 30. The step for determining and storing at least one predictive model is performed by a predictive model determining circuit **50** as described on page 20, lines 20-33.

The invention of claim 2 is directed to the computer-implemented method of claim 1, further including, as shown in FIG. 2, determining an inter-story similarity metric for the story-pairs, and determining source-pair statistics for the story-pairs for the inter-story similarity vectors recited in claim 1.

The invention of claim 20 is directed to a linked event detection training system that includes an input/output circuit **10** (FIG. 4), a memory **20**, and a processor **15**, as shown in FIG. 4, that receives source-identified training stories and associated link label information for at least one story-pair via the input/output circuit. The link label information indicates the existence of at least one link between a pair of stories in the source-identified training stories and that the linked source-identified stories are related to the same event. Also included in the system is an inter-story similarity vector determining circuit **40, 45** that determines inter-story similarity vectors in the memory for at least one story-pair of the source-identified training stories. Further included is a predictive model determining circuit **50** that determines and stores at least one predictive model in the memory based on the inter-story similarity vectors and the link label information.

The invention of claim 21 is directed to the linked event detection training system of claim 20, further including, as shown in FIG. 2, a similarity metric determining circuit that determines an inter-story similarity metric for the story-pairs, and a similarity statistics determining circuit that determines source-pair statistics for the story-pairs for the inter-story similarity vectors recited in claim 20.

The invention of claim 39 is directed to a computer-implemented method of linked event detection as shown in FIG. 3, including: determining source-identified training stories, determining inter-story similarity vectors in a memory **20** (FIG. 4) for at least one story-pair of the source-identified training stories, determining at least one predictive model in the memory for link detection, determining a link between the story-pairs based on the predictive model and the inter-story similarity vector, and displaying the link on a computer **300** or storing the link in an information repository **200**. The link indicates that the story-pairs are related to the same event.



A processor **15**, as shown in FIG. 4 of the specification, performs the step of determining source-identified training stories as described on page 20, line 20 – page 21, line 3. The step for determining inter-story similarity vectors is performed by an inter-story similarity determining circuit **40** as described on page 21, line 28 – page 22, line 30. The step for determining at least one predictive model is performed by a predictive model determining circuit **50** as described on page 20, lines 20-33. The step for determining a link between the story-pairs is performed by a link determining circuit **55** as described on page 22, lines 20-30. The step for displaying the link on a computer or storing the link in an information repository is performed by the link determining circuit.

The invention of claim 40 is directed to the computer-implemented method of claim 39, further including, as shown in FIG. 3, determining an inter-story similarity metric for each story-pair, and determining source-pair statistics for the story-pairs for the inter-story similarity vectors recited in claim 39.

The invention of claim 58 is directed to a linked event detection system that includes an input/output circuit **10** (FIG. 4), a memory **20**, and a processor **15**, as shown in FIG. 4, that receives source-identified stories via the input/output circuit. An inter-story similarity vector determining circuit **40, 45** determines inter-story similarity vectors in the memory for the story-pairs of the source-identified stories. A link determining circuit **55** determines and displays on a computer **300** or stores in an information repository **200**, links between story-pairs based on a predictive model in the memory and the inter-story similarity vectors. The links indicate that the story-pairs are related to the same event.

The invention of claim 59 is directed to the linked event detection system of claim 58, further including, as shown in FIG. 3, a similarity metric determining circuit that determines an inter-story similarity metric for the story-pairs, and a similarity statistics

determining circuit that determines source-pair statistics for the story-pairs for the inter-story similarity vectors recited in claim 58.

The invention of claim 77 is directed to a method of determining a stopwords list as shown in FIG. 5, including: determining a source-identified training corpus of text information, determining a verified first source-mode transformation of the source-identified training corpus text from a first mode to a second mode based on a verified transcription or a verified translation, determining an un-verified second source-mode transformation of the source-identified training corpus text from a first mode to a second mode, determining at least one transformation error associated with distribution differences between the first and second transformations and identified sources, determining and storing at least one source-specific transformation action for the determined transformation errors in a memory **20** (FIG. 4), and identifying and transforming transformation errors in other transformed source-identified texts based on the source-specific transformation actions in the memory.

A processor **15**, as shown in FIG. 4 of the specification, performs the steps of determining a source-identified training corpus, determining a verified first source-mode transformation, determining an un-verified second source-mode transformation, determining at least one transformation error, determining and storing at least one source-specific transformation action, and identifying and transforming transformation errors in other transformed source-identified texts based on the source-specific transformation actions in the memory.

The invention of claim 81 is directed to computer readable storage medium comprising computer readable program code embodied on the computer readable storage medium. The computer readable program code is processable to program a computer **15** to determine at least one predictive model for a linked event detection

system **100** by executing the following recited steps. Claim 81 recites steps, as shown in FIG. 2, for determining source-identified training stories, determining inter-story similarity vectors in a memory **20** (FIG. 4) for at least one story-pair, determining link label information for the at least one story-pair, and determining and storing at least one predictive model in the memory based on the inter-story similarity vectors and the link label information. The link label information indicates the existence of at least one link between a pair of stories in the source-identified training stories and that the linked source-identified stories are related to the same event.

A processor **15**, as shown in FIG. 4 of the specification, performs the step of determining source-identified training stories and the step of determining link label information as described on page 20, line 20 – page 21, line 3. The step for determining inter-story similarity vectors is performed by an inter-story similarity determining circuit **40** as described on page 21, line 28 – page 22, line 30. The step for determining and storing at least one predictive model is performed by a predictive model determining circuit **50** as described on page 20, lines 20-33.

The invention of claim 82 is directed to computer readable storage medium comprising computer readable program code embodied on the computer readable storage medium. The computer readable program code is processable to program a computer **15** to determine at least one predictive model for a linked event detection system **100**. Each of the instructions limitations in claim 82 recites a §112, 6<sup>th</sup> paragraph, means-plus-function limitation and the disclosed structures, materials, or acts described in the specification that correspond to the claimed step are identified with reference to FIG. 4. The computer readable program code includes instructions for determining source-identified training stories (processor **15**, page 20, line 20 – page 21,

line 3), instructions for determining inter-story similarity vectors in a memory **20** for at least one story-pair (circuit **40**, page 21, line 28 – page 22, line 30), instructions for determining link label information (processor **15**, page 20, line 20 – page 21, line 3) for the at least one story-pair, and instructions for determining and storing at least one predictive model in the memory based on the inter-story similarity vectors and the link label information (circuit **50**, page 20, lines 20-33). The link label information indicates the existence of at least one link between a pair of stories in the source-identified training stories and that the linked source-identified stories are related to the same event.

The invention of claim 83 is directed to computer readable storage medium comprising computer readable program code embodied on the computer readable storage medium. The computer readable program code is processable to program a computer **15** to detect linked events by executing program steps. Claim 83 recites steps, as shown in FIG. 3, for determining source-identified training stories, determining inter-story similarity vectors in a memory **20** (FIG. 4) for at least one story-pair of the source-identified training stories, determining at least one predictive model in the memory for link detection, determining a link between the story-pairs based on the predictive model and the inter-story similarity vectors, and displaying the link on a computer **300** or storing the link in an information repository **200**. The link indicates that the story-pairs are related to the same event.

A processor **15**, as shown in FIG. 4 of the specification, performs the step of determining source-identified training stories as described on page 20, line 20 – page 21, line 3. The step for determining inter-story similarity vectors is performed by an inter-story similarity determining circuit **40** as described on page 21, line 28 – page 22, line

30. The step for determining at least one predictive model is performed by a predictive model determining circuit **50** as described on page 20, lines 20-33. The step for determining a link between the story-pairs is performed by a link determining circuit **55** as described on page 22, lines 20-30. The step for displaying the link on a computer or storing the link in an information repository is performed by the link determining circuit.

The invention of claim 84 is directed to directed to computer readable storage medium comprising computer readable program code embodied on the computer readable storage medium. The computer readable program code is processable to program a computer **15** to detect linked events. Each of the instructions limitations in claim 84 recites a §112, 6<sup>th</sup> paragraph, means-plus-function limitation and the disclosed structures, materials, or acts described in the specification that correspond to the claimed step are identified with reference to FIG. 4. The computer readable program code includes instructions for determining source-identified training stories (processor **15**, page 20, line 20 – page 21, line 3), instructions for determining inter-story similarity vectors in a memory **20** for at least one story-pair of the source-identified training stories (circuit **40**, page 21, line 28 – page 22, line 30), instructions for determining at least one predictive model in the memory for link detection (circuit **50**, page 20, lines 20-33), instructions for determining a link between the story-pairs based on the predictive model and the inter-story similarity vectors (link determining circuit **55**, page 22, lines 20-30), and instructions for displaying the link on a computer **300** or storing the link in an information repository **200** (performed by the link determining circuit **55**). The link indicates that the story-pairs are related to the same event.

VI. GROUNDS OF REJECTION TO BE REVIEWED ON APPEAL

The following grounds of rejection are presented for review:

1) Whether claims 1-2, 20-21, 39-40, 58-59, and 81-84 are unpatentable under 35 U.S.C. §103(a) over U.S. Patent No. 6,606,620 issued to Sundaresan et al. (hereinafter Sundaresan) in view of U.S. Patent No. 5,835,905 issued to Pirolli et al. (hereinafter Pirolli) and further in view of U.S. Patent No. 6,961,954 issued to Maybury et al. (hereinafter Maybury).

2) Whether claim 77 is unpatentable under 35 U.S.C. §103(a) over Brown, Ralf D. "Dynamic Stopwording for Story Link Detection" (hereinafter Brown) in view of U.S. Patent No. 6,012,073 issued to Arend et al. (hereinafter Arend).

## VII. ARGUMENT

### A. Claims 1-2, 20-21, 39-40, 58-59, and 81-84 Would Not Have Been Obvious Over Sundaresan in View of Pirolli and Further in View of Maybury

#### 1. Claim 1

##### a. Sundaresan Does Not Teach Use of Source Identification

In the Amendment After Final Rejection filed on January 21, 2008 Applicants argued that the Office Action failed to show where the Sundaresan reference makes any reference to the use of source information for the training documents. It was also argued that the Sundaresan reference appears to be silent on the subject of source information for the training documents, but instead describes only word or term-based vectors. In the Advisory Action mailed February 19, 2008, however, it is argued in the Continuation Sheet that Applicant's argument was found unpersuasive. It was further stated that Sundaresan teaches in col. 6, lines 48-49 that web pages are identified by URL and come from web sites (col. 6, lines 65-68) that are associated with particular domain names and include the content of a particular organization. However, neither the Final Office Action nor the Advisory Action show where Sundaresan describes using the URL or even the domain name in embodiments taught by Sundaresan. To the contrary, the URL and domain name described in col. 6, lines 48-49 and lines 65-67 are only being discussed in a definition of background explanation section (see col. 5, lines 44-47) of the Sundaresan reference and are not further discusses in the reference except to show that the result if a user search on the Internet may include a list of URLs (col. 7, lines 40-42). The Examiner has not shown where the Sundaresan reference describes utilization of the source information for a document. Merely knowing that a

document had a source, or knowing that a document is being retrieved from a source, does not teach or suggest utilization of the source identification as part of determining predictive models for linked event detection. To the contrary, however, the source information of a document is clearly a recited feature in the limitations of claim 1 of the present application, i.e., source-identified.

b. Combination of Sundaresan and Pirolli Does Not Teach Determining Inter-Story Similarity Vectors for Source-Identified Training Stories

The Final Office Action admits that Sundaresan does not teach determining inter-story similarity vectors for at least one story-pair, however, the Office Action argues that Pirolli teaches determining inter-story similarity vectors, with reference to col. 7, lines 53-65. However, as argued in the Amendment, the vectors taught by Pirolli are clearly word-based and do not include source information. The Advisory Action, however, in the Continuation Sheet, argues that because Sundaresan teaches source, that the broadest interpretation of the limitation is covered by combining the references. Applicants submit that the Examiner is reading features into the references that are neither taught nor suggested as described above under "Sundaresan Does Not Teach Use of Source Identification." The limitations of claim 1 clearly recite determining inter-story similarity vectors for source-identified training stories.

c. Sundaresan Does Not Teach Determining Link label Information for Source-Identified Training Stories

Applicants argued in the Amendment After Final Rejection that Sundaresan does not teach determining link label information. The Advisory Action asserts that this argument is not persuasive and that classifying documents into categories means that



the documents in a particular directory are linked by similarity of terms or concept. However, classifying documents into directories based on similarity of terms or concept is a much broader categorization than determining that two stories are linked to the same event. The Examiner has not shown where the Sundaresan reference teaches or describes determining that stories are related to the same event. Claim 1, however, clearly recites a limitation for determining link-label information that indicates that the source-identified stories are related to the same event.

d. Maybury Does Not Teach Indicating Existence  
of Stories Linked to the Same Event

Applicants argued in the Amendment After Final Rejection that Maybury does not teach indicating the existence of stories linked to the same event. The Advisory Action asserts that this argument is not persuasive and asserts that Maybury teaches a system that finds interrelated stories using segmentation (col. 19, lines 33-38). Contrary to the Examiner's interpretation of Maybury, i.e., finding interrelated stories using segmentation, the Maybury reference teaches segmenting a given story into segments for, e.g., more timely and efficient communication and storage of multimedia data (col. 2, lines 41-67). In fact, col. 19, lines 33-38 cited by the Examiner teach exactly the opposite of finding stories linked to the same event. The cited text describes how the "system over-generated story segments, which increased the number of stories that were found. In three cases of over segmentation, a story crossed a commercial boundary and was broken into two individual stories. In one case, a story consisted of two sub-stories, the Peruvian Army and the Peruvian Economy." As evidenced by this cited section of the Maybury reference, it seems the goal is to separate an individual document into multiple documents based on the subject matter, rather than finding

separate stories linked to the same event. As described above, claim 1 clearly recites a limitation for determining link label information indicating a link between a pair of stories, and that the linked source-identified stories are related to the same event.

2. Claims 20, 39, 58 and 81-84

Each of claims 20, 39, 58 and 81-84 recites limitations similar to the above-discussed limitations of claim 1. Therefore, each of the arguments discussed above with reference to claim 1 apply as well to claims 20, 39, 58 and 81-84. Consequently, Applicants direct the Board's attention to the arguments for claim 1 and have not repeated the arguments here.

3. Claim 2

a. Sundaresan Does Not Teach Use of Source Identification

Applicants note that the Final Office Action mailed September 14, 2007 admits that Sundaresan does not teach determining inter-story similarity vectors for at least one story-pair. The Office Action then states that Pirolli teaches determining inter-story similarity vectors, with reference to col. 7, lines 53-65. However, the vectors taught by Pirolli are clearly word-based and do not include source information. For example, Pirolli teaches that the "token information is then used to create a document vector, where each component of the vector represents a word, step 403. Entries in the vector for a document indicate the presence or frequency of a word in the document. The steps 401-403 are repeated for each Web page in the Web locality. For each pair of pages, the dot product of these vectors is computed, step 404. The dot product which produces a similarity measure" (col. 7, lines 58-65, underlining added).

Contrariwise, claim 2 of the present application recites limitations wherein the recited inter-story similarity vectors include two components: at least one inter-story similarity metric and at least one source-pair statistics. FIG. 6 of the present application shows the steps of determining source-pair similarity statistics which comprise a component of the inter-story similarity vectors. The process is further described on page 24, line 21 - page 25, line 26 of the present application. It should be noted that in step S1030, the source pair statistics are determined based on the source characteristics of the stories in the source-pair. Specific source pair statistics are maintained for each identified source pair.

Unlike the document vector of Pirolli, the recited source pair statistics are not document/story word or term based. As described on page 24, line 29 – page 25, line 11, the source characteristics upon which the recited source pair statistics are based are associated with a source which may be a CNN, ABC, NBC, Aljazeera or CTV television broadcast, the text of a Reuters newswire service story, an article in the Wall Street Journal or any other known or later developed information source. The source characteristics associated with each source in a source-pair are used to select source-pair similarity statistics from the source hierarchy. The source hierarchy may be based on source characteristics such as source language, input mode and the like. An English radio broadcast captured using automatic speech recognition may be associated with an "English" language source characteristic and an "ASR" input mode source characteristic. A Chinese text translated into English may be associated with a "Chinese" source language characteristic and a "TEXT" input mode characteristic. The two stories thus form a story pair having "English:ASR" and "Chinese:TEXT" source pair characteristics.

Use of the above-described source-pair statistics as a component of similarity vectors, in combination with inter-story similarity vectors, as recited in claim 2, are neither taught nor suggested in the cited references which describe only word/term-based vectors. Thus, the cited references, either individually or combined, do not teach the inter-story similarity vectors recited in claim 2. Applicants also note that the Office Action cites col. 10, lines 15-17 of Sundaresan as teaching "determining at least one source-pair statistics" as recited in claim 2. However, as described above, the statistics computed in the Sundaresan reference are term-based, and source information is not discussed. This is further made clear in lines 17-22 of col. 10: "The statistics are calculated by combining all the documents of a given type together in a meaningful fashion. In particular, the modeling sub-module 415 combines the individual vectors in the class by adding them together and normalizing the result. Term frequencies may be normalized at any level from the uppermost (document level) to the lowest sub-vector." The described step of adding individual vectors is merely adding previously determined term-based vectors and normalizing the results.

4. Claims 21, 40, and 59

Each of claims 21, 40, and 59 recites limitations similar to the above-discussed limitations of claim 2. Therefore, each of the arguments discussed above with reference to claim 2 apply as well to claims 21, 40, and 59. Consequently, Applicants direct the Board's attention to the arguments for claim 2 and have not repeated the arguments here.

B. Claim 77 Would Not Have Been Obvious Over Brown in View of Arend

Applicants argued in the Amendment After Final Rejection that Brown does not teach determining a stopword list based on first and second source-mode transformations, and determining a transformation error associated with distribution differences between the first and second transformations. The Advisory Action argues that Brown describes transformation by removing stopwords from the text. Applicants submit, however that Brown does not teach or suggest the above-mentioned first and second source-mode transformations, and determining a transformation error associated with distribution differences between the first and second transformations. However, in the Continuation Sheet of the Advisory Action, the following is argued:

As to verified and unverified the specification paragraphs 0115 and 0116 [paragraphs 0086 and 0087 as filed] of instant application are confusing to interpret. The specification seems to point to translating a story using "trusted" translation and parts of the text that are not transformed are then used for second transformation. The claimed limitation seems to direct to two separate transformations of the same text. The specification also gives exemplary transformation in paragraph 0035 [paragraph 0021 as filed] as being "performed manually several time until a certain result is agreed as correct". As such the process seems to be performed by a person, which cannot be patented.

The paragraphs referenced above in the Advisory Action are quoted below in the order referenced for convenience:

[0115] A verified first transformation of the corpus from a first mode to a second mode is determined in step S530. For example, in one exemplary embodiment according to this invention, verified transcriptions of the training corpus from speech utterances to text using automatic speech recognition are determined. It will be apparent that other

transformations such as language translations may be similarly verified using known translations such as the United Nations official reports, Hansard's official English/French translations of Canadian Parliamentary Reports and/or any known or later developed parallel corpus of verified transformed text. Once the verified first transformation is determined, control continues to step S540.

[0116] In step S540, an un-verified second transformation of the corpus from a first mode to a second mode is determined. The un-verified second transformation reflects the actual use of the transformation within a process or system. The errors induced by the transformation are reflected in the un-verified second transformation of the corpus. After the un-verified second transformation is determined, control continues to step S550.

[0035] A verified transformation is a transformation that has been checked for accuracy. For example, the transformation may be performed manually several times and a certain result agreed as correct. In various exemplary embodiments, a standard parallel translation corpus such as United Nations Reports, Hansard's Parliamentary Reports, and the like are used to determine a verified translation. An un-verified translation is determined by applying the translation process to each of the un-transformed texts in the corpus. The differences between the verified translation and the un-verified translation reflect errors induced by the translation and/or transformation process. It will be apparent that in various other exemplary embodiments according to this invention, any known or later developed method of determining a verified transformation and/or any other method of determining systematic errors induced by the transformation process may be used in the practice of this invention.

First, with reference with the transformation process seemingly being performed by a person, Applicants submit that the Examiner has misinterpreted the specification. While the specification states: "For example, the transformation may be performed manually several times and a certain result agreed as correct", the specification is

clearly not describing a manual transformation process, only that the process may be invoked manually several times. The transformation itself is not being performed by a human, but rather by the processor **15**, as clearly stated on page 21, lines 4-5, of the specification as filed.

Secondly, with reference to the portion of the specification which the Advisory Action describes as confusing, i.e., the Advisory Action asserts that the specification seems to point to translating a story using “trusted” translation and parts of the text that are not transformed are then used for second transformation while, on the other hand, the claimed limitation seems to direct to two separate transformations of the same text. Applicants respectfully submit that in both the specification and the recited claim limitation, two separate transformations are performed on the same text. On page 23, lines 11-12, the specification describes a “verified first transformation of the corpus from a first mode to a second mode is determined in step S530.” On page 23, lines 20-21, the specification describes determining “an un-verified second transformation of the corpus from a first mode to a second mode.” In both cases, it is the corpus that is being transformed. i.e., two separate transformations of the same text. The second transformation is not described as being performed on “parts of the text that are not transformed” as asserted in the Advisory Action. Neither the Final Office Action nor the Advisory Action show where Brown teaches this recited feature of claim 77.

### CONCLUSION

For all of the reasons discussed above, it is respectfully submitted that the rejections are in error and that independent claims 1, 20, 39, 58, 77, and 81-84 are in

condition for allowance. Applicants submit also that each of the remaining dependent claims 2-19, 21-38, 40-57, 59-76, 78-80, and 85-88, by reason of dependence from their base independent claims, are also in condition for allowance. For all of the above reasons, Appellants respectfully request this Honorable Board to reverse the rejections of claims 1-88.

Respectfully submitted,

A handwritten signature in black ink, appearing to read 'M. Svat', is written over a horizontal line.

Mark S. Svat  
Registration No. 34,261

MSS:TGY/kan

FAY SHARPE LLP  
1100 Superior Avenue – Seventh Floor  
Cleveland, Ohio 44114-2579  
Telephone: (216) 861-5582

Filed: April 14, 2008

N:\XERZ\201564\GTY0000452V001.doc



## **APPENDICES**

### **VIII. CLAIMS APPENDIX:**

Claims involved in the Appeal are as follows:

1. A computer-implemented method of determining predictive models for a linked event detection system comprising the steps of:

determining source-identified training stories;

determining inter-story similarity vectors in a memory for at least one story-pair of the source-identified training stories;

determining link label information for the at least one story-pair, the link label information indicating the existence of at least one link between a pair of stories in the source-identified training stories and that the linked source-identified stories are related to the same event; and

determining and storing at least one predictive model in the memory based on the inter-story similarity vectors and the link label information.

2. The method of claim 1, wherein the step of determining inter-story similarity vectors comprises the steps of:

determining at least one inter-story similarity metric for the story-pairs; and

determining at least one source-pair statistics for the at least one story-pair.

3. The method of claim 2, wherein determining inter-story similarity vectors further comprise the step of normalizing the inter-story similarity metric based on the source-pair statistics.

4. The method of claim 2, wherein determining inter-story similarity vectors further comprise the step of incrementally normalizing the inter-story similarity metric based on the source-pair statistics.

5. The method of claim 2, wherein the inter-story similarity metric is normalized based on at least one of subtraction and division.

6. The method of claim 2, wherein the inter-story similarity metric is at least one of a probability based similarity metric and a Euclidean based similarity metric.

7. The method of claim 6, wherein the probability based inter-story similarity metric is at least one of a Hellinger, a Tanimoto and a clarity distance based metric.

8. The method of claim 6, wherein the Euclidean based inter-story similarity metric is a cosine-distance based metric.

9. The method of claim 1, further comprising the step of transforming the source-identified training stories.

10. The method of claim 9, wherein transforming the source-identified training stories is at least one of translating, transcribing and linguistically transforming.

11. The method of claim 2, wherein the inter-story similarity metrics are based on terms in at least one source-identified term frequency-inverse story frequency models.

12. The method of claim 11, wherein the terms in source-identified term frequency-inverse story frequency models are based on language.

13. The method of claim 11, wherein determining terms comprises the steps:  
determining a reference language; and  
determining reference language and non-reference language terms.

14. The method of claim 2, wherein the at least one inter-story similarity metric is normalized based on at least one of a source-pair identified similarity statistic.

15. The method of claim 1, wherein the at least one predictive model is at least one of: a classifier, a support vector machine, a decision tree and a Naive-Bayes classifier.

16. The method of claim 2, wherein at least one of the source-pair similarity statistics are determined based on a source hierarchy.

17. The method of claim 16 wherein the source hierarchy is determined based on at least one source characteristic.

18. The method of claim 16 wherein the source characteristic is at least one of a language characteristic, an input mode characteristic, a genre characteristic, a source name characteristic and a transformation characteristic.

19. The method of claim 16 wherein the source-pair similarity statistic for a new source is determined based on at least one source characteristic of the new source.

20. A linked event detection training system comprising:

an input/output circuit;

a memory;

a processor that receives source-identified training stories and associated link label information for at least one story-pair via the input/output circuit, the link label information indicating the existence of at least one link between a pair of stories in the source-identified training stories and that the linked source-identified stories are related to the same event;

an inter-story similarity vector determining circuit that determines inter-story similarity vectors in the memory for at least one story-pair of the source-identified training stories; and

a predictive model determining circuit that determines and stores at least one predictive model in the memory based on the inter-story similarity vectors and the link label information.

21. The system of claim 20, wherein the inter-story similarity vector determining circuit is comprised of:

a similarity metric determining circuit that determines at least one inter-story similarity metric for the at least one story-pair; and

a similarity statistics determining circuit that determines at least one source-pair statistic for the at least one story-pair.

22. The system of claim 21, wherein the inter-story similarity vector determining circuit normalizes the inter-story similarity metric based on the source-pair statistics.

23. The system of claim 21, wherein the inter-story similarity vector determining circuit incrementally normalizes the inter-story similarity metric based on the source-pair statistics.

24. The system of claim 21, wherein at least one of the inter-story similarity metrics is normalized based on at least one of a subtraction and a division operation.

25. The system of claim 21, wherein at least one of the inter-story similarity metrics is at least one of a probability based similarity metric and a Euclidean based similarity metric.

26. The system of claim 25, wherein the probability based inter-story similarity metric is at least one of a Hellinger, a Tanimoto and a clarity distance based metric.

27. The system of claim 25, wherein the Euclidean based inter-story similarity metric is a cosine-distance based metric.

28. The system of claim 20, wherein the source-identified training stories are transformed.

29. The system of claim 28, wherein transforming the source-identified training stories is at least one of translating, transcribing and linguistically transforming.

30. The system of claim 20, wherein the inter-story similarity metrics are based on terms in at least one source-identified term frequency-inverse story frequency model.

31. The system of claim 30, wherein the terms in the source-identified term frequency-inverse story frequency models are based on language.

32. The system of claim 30, wherein the processor determines terms based on a reference language; and determining reference language and non-reference language terms.

33. The system of claim 21 wherein the at least one inter-story similarity metric is normalized based on at least one of a source-pair identified similarity statistic.

34. The system of claim 20, wherein the at least one predictive model is at least one of: a classifier, a support vector machine, a decision tree and a Naive-Bayes classifier.

35. The system of claim 21, wherein the source-pair identified similarity statistic is determined based on a source hierarchy.

36. The system of claim 35, wherein the source hierarchy is determined based on at least one of a source characteristic.

37. The system of claim 35, wherein the source characteristic is at least one of a language characteristic, an input mode characteristic, a genre characteristic, a source name characteristic and a transformation characteristic.

38. The system of claim 35, wherein the source-pair similarity statistic for a new source is determined based on at least one source characteristics of the new source.

39. A computer-implemented method of linked event detection comprising the steps of:

determining source-identified stories;

determining inter-story similarity vectors in a memory for the story-pairs of the source-identified stories;

determining at least one predictive model in the memory for link detection;

determining a link between the story-pairs based on the predictive model and the inter-story similarity vector; and

displaying the link on a computer or storing the link in an information repository, the link indicating the story-pairs are related to the same event.

40. The method of claim 39, wherein the step of determining inter-story similarity vectors comprises the steps of:

determining at least one inter-story similarity metric for each story-pair; and

determining source-pair statistics for the story-pairs.

41. The method of claim 40, wherein determining inter-story similarity vectors further comprise the step of normalizing the inter-story similarity metric based on the source-pair statistics.

42. The method of claim 40, wherein determining inter-story similarity vectors further comprise the step of incrementally normalizing the inter-story similarity metric based on the source-pair statistics.

43. The method of claim 40, wherein the inter-story similarity metric is normalized based on at least one of subtraction and division.

44. The method of claim 40, wherein the inter-story similarity metric is at least one of a probability based similarity metric and a Euclidean based similarity metric.

45. The method of claim 44, wherein the probability based inter-story similarity metric is at least one of a Hellinger, a Tanimoto and a clarity distance based metric.

46. The method of claim 44, wherein the Euclidean based similarity metric is a cosine-distance based metric.

47. The method of claim 39, further comprising the step of transforming the source-identified training stories.

48. The method of claim 47, wherein transforming the source-identified training stories is at least one of translating, transcribing and linguistically transforming.

49. The method of claim 40, wherein the inter-story similarity metrics are based on terms in at least one source-identified term frequency-inverse story frequency models.



50. The method of claim 49, wherein the terms in source-identified term frequency-inverse story frequency models are based on language.

51. The method of claim 49, wherein determining terms comprises the steps:  
determining a reference language; and  
determining reference language and non-reference language terms.

52. The method of claim 40, wherein the at least one inter-story similarity metric is normalized based on at least one of a source-pair identified similarity statistic.

53. The method of claim 39, wherein the at least one predictive model is at least one of: a classifier, a support vector machine and a decision tree, a Naive-Bayes-classifier.

54. The method of claim 40, wherein the source-pair identified similarity statistic is determined based on a source hierarchy.

55. The method of claim 54, wherein the source hierarchy is determined based on at least one of a source characteristic.

56. The method of claim 54, wherein the source characteristic is at least one of a language characteristic, an input mode characteristic, a genre characteristic, a source name characteristic and a transformation characteristic.

57. The method of claim 54, wherein the source-pair similarity statistic for a new source is determined based on at least one source characteristics of the new source.

58. A linked event detection system comprising:

an input/output circuit;

a memory;

a processor that receives source-identified stories via the input/output circuit;

an inter-story similarity vector determining circuit that determines inter-story similarity vectors in the memory for the story-pairs of the source-identified stories; and

a link determining circuit that determines and displays on a computer or stores in an information repository, links between story-pairs based on a predictive model in the memory and the inter-story similarity vectors, the links indicating the story-pairs are related to the same event.

59. The system of claim 58, wherein the inter-story similarity vector determining circuit is comprised of:

a similarity metric determining circuit that determines at least one inter-story similarity metric for the story-pairs; and

a similarity statistics determining circuit that determines source-pair statistics for the story-pairs.

60. The system of claim 59, wherein the inter-story similarity vector determining circuit normalizes the inter-story similarity metric based on the source-pair statistics.

61. The system of claim 59, wherein the inter-story similarity vector determining circuit incrementally normalizes the inter-story similarity metric based on the source-pair statistics.

62. The system of claim 59, wherein at least one of the inter-story similarity metrics is normalized based on at least one of a subtraction and a division operation.

63. The system of claim 59, wherein at least one of the inter-story similarity metrics is at least one of a probability based similarity metric and a Euclidean based similarity metric.

64. The system of claim 63, wherein the probability based inter-story similarity metric is at least one of a Hellinger, a Tanimoto and a clarity distance based metric.

65. The system of claim 63, wherein the Euclidean based inter-story similarity metric is a cosine-distance based metric.

66. The system of claim 58, wherein the source-identified training stories are transformed.

67. The system of claim 66, wherein transforming the source-identified training stories is at least one of translating, transcribing and linguistically transforming.

68. The system of claim 59, wherein the inter-story similarity metrics are based on terms in at least one source-identified term frequency-inverse story frequency model.

69. The system of claim 68, wherein the terms in the source-identified term frequency-inverse story frequency models are based on language.

70. The system of claim 68, wherein the processor determines terms based on a reference language; and non-reference language terms.

71. The system of claim 59, wherein the at least one inter-story similarity metric is normalized based on at least one of a source-pair identified similarity statistic.

72. The system of claim 58, wherein the predictive model is at least one of: a classifier, a support vector machine and a decision tree, a Naive-Bayes classifier.

73. The system of claim 59, wherein the source-pair identified similarity statistic is determined based on a source hierarchy.

74. The system of claim 73, wherein the source hierarchy is determined based on at least one of a source characteristic.

75. The system of claim 73, wherein the source characteristic is at least one of a language characteristic, an input mode characteristic, a genre characteristic, a source name characteristic and a transformation characteristic.

76. The system of claim 73, wherein the source-pair similarity statistic for a new source is determined based on at least one source characteristics of the new source.

77. A method of determining a stopwords list comprising the steps of:

determining a source-identified training corpus of text information;

determining a verified first source-mode transformation of the source-identified training corpus text from a first mode to a second mode based on at least one of a verified transcription and a verified translation;

determining an un-verified second source-mode transformation of the source-identified training corpus text from a first mode to a second mode;

determining at least one transformation error associated with distribution differences between the first and second transformations and identified sources;

determining and storing at least one source-specific transformation action for the determined transformation errors in a memory; and

identifying and transforming transformation errors in other transformed source-identified texts based on the source-specific transformation actions in the memory.

78. The method of claim 77, wherein the first mode is at least one of a text source, an optical character recognition source and an automatic speech recognition source.

79. The method of claim 77, wherein the second mode is at least one of a text source, an optical character recognition source and an automatic speech recognition source.

80. The method of claim 77, wherein the source-specific transformation is at least one of a removal, a repair and a normalization transformation.

81. Computer readable storage medium comprising: computer readable program code embodied on the computer readable storage medium, the computer readable program code processable to program a computer to determine at least one predictive model for a linked event detection system by executing steps comprising:

determining source-identified training stories;

determining inter-story similarity vectors in a memory for at least one story-pair;

determining link label information for the at least one story-pair of the source-identified training stories, the link label information indicating training stories related to the same event; and

determining and storing at least one predictive model in the memory based on the inter-story similarity vectors and the link label information.

82. Computer readable storage medium comprising: computer readable program code embodied on the computer readable storage medium, the computer readable program code processable to program a computer to determine at least one predictive model for a linked event detection system, the computer readable program code comprising:

instructions to determine source-identified training stories;

instructions to determine inter-story similarity vectors in a memory for at least one story-pair of the source-identified training stories;

instructions to determine link label information for the at least one story-pair, the link label information indicating training stories related to the same event; and instructions to determine and store at least one predictive model in the memory based on the inter-story similarity vectors and the link label information.

83. Computer readable storage medium comprising: computer readable program code embodied on the computer readable storage medium, the computer readable program code processable to program a computer to detect linked events by executing steps comprising :

determining source-identified stories;

determining inter-story similarity vectors in a memory for the at least one story-pair of the source-identified stories;

determining at least one predictive model in the memory for link detection; and

determining a link between story-pairs based on the at least one predictive model and the inter-story similarity vectors, the link indicating the story-pairs are related to the same event; and

displaying the link on a computer or storing the link in an information repository.

84. Computer readable storage medium comprising: computer readable program code embodied on the computer readable storage medium, the computer readable program code processable to program a computer to detect linked events, the computer readable program code comprising:

instructions to determine source-identified stories;

instructions to determine inter-story similarity vectors in a memory for the at least one story-pair of the source-identified stories;

instructions to determine at least one predictive model in the memory for link detection;

instructions to determine a link between story-pairs based on the predictive model and the inter-story similarity vectors, the link indicating the story-pairs are related to the same event; and

instructions to display the link on a computer or store the link in an information repository.

85. The method of claim 2, wherein determining at least one source-pair statistic for the at least one story-pair is based on at least one of a similarity metric and a statistic associated with the metric.

86. The system of claim 21, wherein determining at least one source-pair statistic for the at least one story-pair is based on at least one of a similarity metric and a statistic associated with the metric.

87. The method of claim 39, wherein at least one of the predictive models is a trained predictive model.

88. The system of claim 58, wherein at least one of the predictive models is a trained predictive model.



IX. EVIDENCE APPENDIX

A copy of each of the following items of evidence relied on by the Appellant  
[and/or the Examiner] is attached:

NONE

X. RELATED PROCEEDINGS APPENDIX  
NONE